Heredity (2003) 91, 322–330

www.nature.com/hdy

Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar

JC Motamayor[1], AM Risterucci[1], M Heath[2] and C Lanaud[1]
[1]*CIRAD, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, TA 40/03, Av. Agropolis, 34398 Montpellier cedex 5, France;* [2]*USDA-ARS, 13601 Old Cutler Road, Miami, FL 33158, USA*

Cacao (*Theobroma cacao* L.) has been cultivated in Central America since pre-Columbian times. The type of cacao cultivated in this region was called Criollo; cacao populations from the Amazon basin were called Forastero. The type of Forastero most commonly cultivated until 1950 was named Amelonado. Historical data show Trinitario cacao to have originated in Trinidad, resulting from natural hybridisation between Criollo and Amelonado Forastero. Doubts persist on the source of the Amelonado Forastero involved in the origin of Trinitario; the Amelonado parent may have come from the Lower Amazon, the Orinoco or the Guyanas. Most of the cacao cultivated worldwide until 1950 consisted of Criollo, Trinitario and Amelonado. From the early 1950s, Forastero material collected in the Upper Amazon region during the 1930s and 1940s began to be employed in breeding programmes. To gain a better understanding of the origin and the genetic basis of the cacao cultivars exploited before the utilisation of germplasm collected in the Upper Amazon, a study was carried out using restriction fragment length polymorphism and microsatellite markers. Trinitario samples from 17 countries were analysed. With molecular markers, it was possible to clearly identify three main genotypes (represented by clones SP1, MAT1-6 and SIAL70) implicated in the origin of most Trinitario clones.
Heredity (2003) **91,** 322–330. doi:10.1038/sj.hdy.6800298

Keywords: *Theobroma cacao* L.; Trinitario; progenitor germplasm; genetic diversity; domestication; molecular markers

Introduction

Theobroma cacao L. ($2n = 2x = 20$) is an important crop in humid tropical regions. Cheesman (1944) proposed two morphogeographic groups within *T. cacao*: Criollo and Forastero. A third group, Trinitario, was described as a hybrid between Criollo and Forastero. Today, the different Forastero populations are studied according to their geographical origins: Upper Amazon, Lower Amazon, Orinoco and Guyana. Forastero trees are also classified according to pod morphology, such as the Amelonado type, characterised by the melonshape of the fruit. Amelonado types are found in Brazil, the Orinoco basin and the Guyanas. In this paper, the term ''cultivar'' will be employed instead of 'group' to refer to the Criollo, Trinitario and Amelonado so-called 'groups', since Motamayor *et al* (2002) showed that these cacao types do not constitute the main genetic groups within the species in the way that was originally proposed (Cheesman, 1944; Cuatrecasas, 1964).

During the pre-Columbian period, Central American civilisations cultivated Criollo cacao. Criollo was also cultivated during the Spanish colonial period in Central and South America (Venezuela and Colombia). Restriction fragment length polymorphisms (RFLP) and microsatellite markers revealed a high level of homozygosity and significantly low genetic diversity within pure Criollo individuals, referred to as Ancient Criollo in Motamayor *et al* (2002). Only eight multilocus genotypes were observed among the 92 Ancient Criollo individuals analysed using 25 RFLP probes. In total, 96% of these 92 individuals shared the same genotype (represented by clone SP1; Motamayor *et al*, 2002). In Ecuador, another cultivar called Nacional was grown which also has a reduced genetic diversity (Lerceteau *et al*, 1997). In Brazil, an Amelonado Forastero type originating from the Lower Amazon was cultivated. Most of the cacao introduced into Africa and Asia during the colonial period originated from Venezuela, Trinidad and Brazil (Wood, 1991), and corresponded to Criollo, Trinitario and Amelonado cultivars, respectively.

Pound (1938, 1945) suggested that Trinitario could have resulted from hybridisation between Criollo and Amelonado from the Guyanas or the Lower Amazon regions. Cheesman (1944) hypothesised that Trinitario originated because of natural hybridisation between the remaining Criollo individuals (after a 'blast' ravaged the Trinidadian Criollo plantations in 1727) and Forastero trees introduced from the Orinoco delta. According to Bartley (personal communication), the formation of hybrid populations between Amazonian types and Criollo in the ancient Criollo plantations could also occur due to introductions into Central America of cultivated cacao from the Lower Amazon.

Trinitario was more productive and resistant to diseases (Pittier, 1935; Cheesman, 1944) compared to Criollo and Amelonado. For this reason, Trinitario became the predominant type; it began to replace Criollo in Central and South America in 1825 (Pittier, 1935) and Nacional in Ecuador in 1890 (Soria, 1970). It also crossed

Correspondence: JC Motamayor, MasterfoodsUSA, 13601 Old Cutler Road, Miami, FL 33158 USA. E-mail: motamayor@hotmail.com
Received 17 July 2002; accepted 25 February 2003

with Amelonado in Brazil and Africa (Silva and Oliveira-Cardoso, 1980; Toxopeus, 1985).

The term 'traditional cultivars' is used in this paper to distinguish Criollo, Amelonado, Trinitario and Nacional from cacao cultivars obtained from the utilisation of Upper Amazon individuals. In 1938, FJ Pound travelled to the Upper Amazon to collect cacao germplasm resistant to Witches' Broom disease (a devastating disease that destroyed the cacao plantings in Ecuador and Trinidad at the beginning of the 20th century and, more recently, those in Brazil). Pound's material was transported to Trinidad and from there it was distributed to most cacao-producing countries. Therefore, the traditional cultivars were the only ones cultivated until approximately 1950, when germplasm from breeding programmes that used Pound's material started being exploited (Lockwood, 1985).

Today, 70% of cacao production is still derived from these traditional cultivars (Paulin and Eskes, 1995), much of which is composed of Trinitario. The genetic structure of Nacional and Criollo traditional cultivars has been characterised (Lerceteau *et al*, 1997; Motamayor *et al*, 2002). The aim of the present study was to characterise the genetic structure of the Trinitario traditional cultivar and to identify the source of the putative Forastero parents, providing new insights into cacao domestication.

Materials and methods

Most of the plant material analysed in this work was previously studied by Motamayor *et al* (2002). In that study, it was demonstrated that most individuals classified as 'Criollo' in germplasm collections (Modern Criollo in Motamayor *et al*, 2002) show introgression of Forastero genes. These hybrid clones were classified as Trinitario in this paper.

Samples of the Trinitario cultivar (137 individuals from 16 countries) were analysed using RFLP markers (Table 1). A representative subsample of 27 Trinitario individuals was studied using microsatellite markers. Nine individuals from Madagascar representing the reference classical Trinitario cultivar were also analysed using microsatellite markers. Indeed, in Madagascar, only the traditional cultivars (Criollo, Trinitario or Amelonado) have been cultivated without other introductions such as Upper Amazon individuals (D Paulin, personal communication).

Criollo individuals, previously defined and analysed by Motamayor *et al* (2002) as Ancient Criollo individuals,

Table 1 List of individuals analysed in the study

Name	*Clt*[a]	*Origin*[b]	*M*[c]	*Name*	*Clt*[a]	*Origin*[b]	*M*[c]	*Name*	*Clt*[a]	*Origin*[b]	*M*[c]
SNK 10	T	Ca	1	LA ESMIDA	T	M	0	ICS 75	T	Td	0
SNK 12	T	Ca	0	LPM2	T	M	0	ICS 89/1	T	Td	1
SNK 64	T	Ca	0	LPM3	T	M	0	ICS 89/2	T	Td	0
SNK 413	T	Ca	0	LPM4	T	M	0	ICS 95	T	Td	1
IFC 6	T	CI	0	LPM6	T	M	0	ALV 1	T	V	0
IFC 7	T	CI	0	PEN16	T	M	1	ALV 4	T	V	0
IFC 11	T	CI	1	RIM 113	T	M	1	ATO5	T	V	0
IFC 19	T	CI	1	RIM-68	T	M	1	CATA 209	T	V	0
IFC 413	T	CI	0	RIM-189	T	M	1	GAL 2	T	V	0
IFC 420	T	CI	0	UIT 5	T	Ma	0	OC 60	T	V	0
IFC 422	T	CI	0	M588	T	Md	2	OC 61	T	V	0
IFC 5	T	CI	1	M589	T	Md	2	OC 63	T	V	0
COL9	T	Col	0	M590	T	Md	2	OC 66	T	V	0
SC 5	T	Col	0	M591	T	Md	2	OC 73	T	V	0
SC 6	T	Col	0	M592	T	Md	2	POR ROJO	T	V	0
SPEC 625	T	Col	0	M593	T	Md	2	PV 2	T	V	0
SPEC 138/8	T	Col	0	M594	T	Md	2	ZEA 206	T	V	0
SPEC 160/9	T	Col	0	M596	T	Md	2	CPC 1	T	V	0
LF 1	T	CR	0	M597	T	Md	2	CRP 2	T	V	0
CRI 216	T	CR	1	CRI 5	T	Nc	1	CS 7	T	V	0
CC 39	T	CR	0	CRI 37	T	Nc	1	JS 202/1	T	V	0
UF 10	T	CR	0	CRI 12	T	Pa	1	JS 202/2	T	V	0
UF 221	T	CR	0	UF 168	T	Pa	1	JS 206	T	V	0
UF 667	T	CR	1	LAFI 7	T	S	0	JS 210	T	V	0
UF 676	T	CR	1	S52	T	ST	0	LMD 1	T	V	0
MOQ 216	T	E	0	ICS 1	T	Td	0	LMD 4	T	V	0
Q 7	T	G	0	ICS 100	T	Td	1	LMD 5	T	V	0
K 5	T	G	0	ICS 39	T	Td	0	LV 1	T	V	0
GS 36	T	Gr	0	ICS 40	T	Td	1	LV 13	T	V	0
DR 1	T	I	1	ICS 48	T	Td	0	LV 14	T	V	0
G 23	T	I	0	ICS 60	T	Td	1	LV 2	T	V	0
WA 40	T	I	0	ICS 84	T	Td	0	LV 3	T	V	0
I059	T	M	0	AC T 2/11	T	Td	0	LV 4	T	V	0
BAN1	T	M	0	ICS 16	T	Td	1	LV 6	T	V	0
OS02	T	M	0	ICS 46	T	Td	0	NR 1	T	V	0
ECH1	T	M	0	ICS 53	T	Td	1	PR01	T	V	0
ECH2	T	M	1	ICS 6	T	Td	0	CEC 2	T	V	0
CHO 28	T	V	0	HE 3	T	V	0	IFC 15	Am	CI	0
CHO 31	T	V	0	HE 5	T	V	0	TJ 1	Am	Hn	0

Table 1 (*continued*)

Name	*Clt*[a]	*Origin*[b]	*M*[c]	*Name*	*Clt*[a]	*Origin*[b]	*M*[c]	*Name*	*Clt*[a]	*Origin*[b]	*M*[c]
CHO 36	T	V	0	HE 6	T	V	0	GU 346	Am	FG	1
CHO 41	T	V	0	HE 2	T	V	0	GU 349	Am	FG	1
CHO 42	T	V	1	HE 201	T	V	0	Atelier 1	Am	Nc	2
CHO 94	T	V	0	VEN 11	Am	OV	1	STA1	Am	Nc	2
CHO 131	T	V	0	VEN 20	Am	OV	1	STA 2	Am	Nc	2
CHU 202	T	V	0	VEN 4	Am	OV	1	UPA 413	UH	CI	0
CHU 24	T	V	0	SIAL 20	Am	B	2	UPA 134	UH	CI	0
ALV 0	T	V	0	SIAL 70	Am	B	1	T 79/416	UH	G	0
CHS 201/12	T	V	0	SIAL 84	Am	B	2	T 63/967	UH	G	0
CHS 205/34	T	V	0	SIAL 169	Am	B	2	T 60/887	UH	G	0
JS 202/1	T	V	0	SIAL 325	Am	B	2	T 85/799	UH	G	0
JS 202/2	T	V	0	SIC 2	Am	B	2	P7XC1	UH	CI	0
JS 211/21	T	V	0	SIC 19	Am	B	2	TSH 1077	UH	Td	0
LV 0	T	V	0	SIC 23	Am	B	2	IFC 1212	UH	CI	0
LV 00	T	V	0	SIC 328	Am	B	2	IFC 1213	UH	CI	0
SAL 2	T	V	0	SIC 662	Am	B	2	SCA6xC1	UH	CI	0
CS 2	T	V	0	SIC 801	Am	B	2	LIB1	C	Nc	0
CS 3	T	V	0	SIC 806	Am	B	2	SP10	C	V	0
CS 5	T	V	0	SIC 864	Am	B	1	SP1	C	V	1
CS 9	T	V	0	NOPT1	Am	B	2	SP2	C	V	0
CATA 211	T	V	0	FCS4	Am	B	2	SP4	C	V	0
CHS 217/18	T	V	0	FSC8	Am	B	2	SP9	C	V	0
CHO 174	T	V	0	FSC15	Am	B	2	ZEA2	C	V	0
CHU 120	T	V	1	MAT 1-6	Am	Cr	1	ZEA3	C	V	0
CHS 205/37	T	V	0	IFC 1	Am	CI	0	LAN24	C	M	1
CS 1	T	V	0	IFC 2	Am	CI	0	LAN30	C	M	0
NOV29	T	V	1	IFC 361	Am	CI	0				
HE 212	T	V	0	IFC 4	Am	CI	1				

[a]Clt = cultivar: C, Criollo; Am, Amelonado Forastero; T, Trinitario; UH, Upper Amazon Forastero hybrids.
[b]Origin = country of origin: B, Brazil; Ca, Cameroon; CI, Côte d'Ivoire; Col, Colombia; CR, Costa Rica; E, Ecuador; FG, French Guiana; G, Ghana; Gr, Grenada; Hn, Honduras; I, Indonesia; Ma, Malaysia; M, Mexico; Md, Madagascar; Nc, Nicaragua; Pa, Panama; S, Samoa; ST, São Tomé; Td, Trinidad; OV, Orinoco (Venezuela); V, Venezuela.
[c]M: 0, individuals analysed with RFLP only; 1, analysed with microsatellite and RFLP markers; 2, analysed with microsatellite only.

consisted of trees sampled from places where gene flow between Criollo and Trinitario or Forastero trees was absent or limited because of the low likelihood of introductions of Trinitario or Forastero material. Criollo constitute the known parent of Trinitario, because, as mentioned above, it was hypothesised that Trinitario arose as a consequence of natural crosses between Criollo and Amelonado Forastero (Pound, 1938, 1945; Cheesman, 1944). In total, 10 Criollo individuals showing the eight genotypes observed through RFLP analysis within this cultivar by Motamayor *et al* (2002) were included in the analysis.

As mentioned above, Forastero trees involved in the origin of Trinitario have three possible geographic origins: the Lower Amazon, the Orinoco basin and the Guyanas. In total, 23 Amelonado Forastero individuals were studied as the Forastero putative parent: 18 from the Lower Amazon (one putatively introduced to the Matina region of Costa Rica and 17 from Brazil), three from the Orinoco, and two from French Guiana. The Amelonado Forastero individuals from Brazil were represented by 13 clones selected by the Agricultural Ministry of Brazil in cacao farms from the state of Bahia (Pinto *et al*, 1967), and four old Amelonado trees sampled from two cacao farms around the city of Itabuna (Fasenda Brasilera and Novo Oriente). Nine individuals, classified as Lower Amazon Amelonado, from germplasm collections of Côte d'Ivoire, Honduras and Nicaragua, were also studied but not included in the parentage analysis because they were suspected of showing introgressions of Criollo alleles. In total, 11 hybrids, with at least one Upper Amazon Forastero parent, were also studied to compare the structure of their genetic diversity to that of Trinitario.

DNA isolation

For RFLP analyses, DNA was purified on caesium chloride gradients as described in Lanaud *et al* (1995). For microsatellite analyses, DNA was isolated according to Risterucci *et al* (2000).

RFLP procedures

All RFLP procedures were conducted as described previously (Lanaud *et al*, 1995). The enzymes used for genomic DNA restriction were *Eco*RI and *Hind*III. In total, 17 cDNA and eight genomic DNA probes, chosen for their coverage of the genetic map of *Theobroma cacao* (Lanaud *et al*, 1995), were used to study 172 individuals of the different types (Table 1).

Microsatellite procedures

In total, 16 microsatellites EMBL accessions—Y16883, Y16977, Y16980, Y16981, Y16983, Y16985, Y16986, Y16988, Y16990, Y16991, Y16992, Y16994, Y16996, Y1697, Y16998 and mtcCIR5 (Lanaud *et al*, 1999)—were used to genotype 50 individuals (Table 1).

Primers were end-labelled with γ^{33}P ATP and amplification was performed in an MJ Research PTC 100 thermal cycler, in 20 µl of reaction mix containing 10 ng cacao DNA, 0.2 mM dNTP mix, 2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.2 µM primer (5′-end labelled with γ^{33}P ATP) and 1 unit of TAQ polymerase (Eurobio). The samples were denatured at 94°C for 4 min, and subjected to 32 repeats of the following cycle: 94°C for 30 s, 46°C or 51°C for 1 min and 72°C for 1 min. After adding 20 µl of loading buffer (98% formamide, 10 mM EDTA, bromophenol blue, xylene cyanol), the mix was denatured at 92°C for 3 min and 3 µl of each sample was loaded in a 5% polyacrylamide gel with 7.5 M urea and electrophoresed in 0.5% TBE buffer at 55 W for 1 h 40 min. The gel was dried for 30 min at 80°C and exposed to X-ray film (Fuji RX) overnight.

Data analyses

RFLP data: Alleles observed after the hybridisation of 25 probes were scored. A factorial analysis of correspondences (FAC) was carried out using the Genetix 4.0 software to study the genetic diversity structure of the Trinitario cultivar.

Microsatellite data: Microsatellite allele sizes were scored by comparison of PCR product lengths to the sequence of the genomic clone from which primers were designed.

Paternity analysis was carried out on a completely heterozygous clone (ICS100) since most Trinitario individuals result from selfing or intercrossing among them. The paternity analysis method, based on calculating a likelihood ratio for each candidate parent (ie the likelihood of parentage of the candidate parent relative to the likelihood of parentage of an arbitrary unrelated candidate parent), was chosen to compare the likelihood ratios of different candidate parents, rather than the exclusion method (Chakraborty *et al*, 1974) in which likelihood ratios cannot be calculated. The likelihood ratio that the observed ICS100 microsatellite alleles were inherited from the known parent (the most frequent Criollo genotype represented by the clone SP1, Motamayor *et al*, 2002) and from each candidate Amelonado parent originating from the three geographical regions was calculated. Only one Criollo individual (SP1) was analysed as the known parent because of the reduced genetic diversity of Criollo (Motamayor *et al*, 2002). SP1 represents the most common genotype, found in 96% of the 92 Ancient Criollo individuals studied by Motamayor *et al* (2002).

The likelihood ratio was calculated using the Cervus 1.0 software (Marshall *et al*, 1998), a simulation programme that generates critical log-likelihood scores to assign paternity at a given level of statistical confidence. Log-likelihood scores are expressed as LOD scores, which are the natural logarithm of the product of the likelihood ratios at each locus. A LOD score is calculated for each candidate parent, based on the genotypes of the candidate parent, putative offspring and known parent (the most frequent Criollo genotype). A negative LOD score implies that the candidate parent is less likely to be the true parent than a randomly chosen genotype. A positive LOD score implies that the candidate parent is more likely to be the true parent than a randomly chosen individual. The most likely candidate parent is the one with the highest LOD score. Delta is the statistic used to assess the reliability of assigning parentage to the most likely candidate parent. Delta is defined as the difference in LOD scores between the most likely candidate parent and the second most likely. The Delta criterion, calculated from the simulations performed from allele frequencies, allows the assessment of the significance of the parentage analysis.

For each Trinitario analysed using microsatellites, proportions of shared alleles (P_{AS}) with the Criollo individual analysed as the known parent of ICS100, and the most likely Forastero parent of ICS100 or genotypes from the same geographical area of the most likely Forastero parent of ICS100 were calculated. $P_{AS} = a/2n$, where a is the number of alleles common to individuals i and j, and n is the number of loci studied. Since P_{AS} between Criollo and Amelonado individuals is zero, each Trinitario will show complementary P_{AS} values totalling 1, resulting from alleles contributed by Criollo and the most likely Forastero parent of ICS100 (or closely related Amelonado genotypes), as long as these are the only genotypes involved in the origin of the Trinitario cultivar. In other words, a given Trinitario will show complementary P_{AS} values totalling 1 with a Criollo and an Amelonado as long as these contain all of the alleles observed in such a Trinitario. For example, a Trinitario (T) having a P_{AS} with a Criollo (C) of 0.4 and a P_{AS} with an Amelonado (A) of 0.6 means that 40% of T alleles are found in C and the remaining 60% in A (knowing that P_{AS} A−C = 0).

The genetic diversity statistic (Nei, 1978), the mean number of microsatellite alleles per locus, the proportion of polymorphic loci when the most frequent allele does not exceed 95% and the observed heterozygosity were calculated for the Lower Amazon Amelonado individuals from Brazil using Genetix 4.0 software. To determine similarity among Amelonado individuals from Brazil, a neighbour-joining tree (Saitou and Nei, 1987) was constructed from the shared allele distance (D_{AS}) between individuals obtained from microsatellite data. The D_{AS} estimation, N–J tree construction and bootstrapping procedures were conducted using a computer program kindly provided by Jean-Marie Cornuet and Sylvain Piry (Laboratoire de Modelisation et Biologie Evolutive, INRA, Montpellier, France).

Results

RFLP results

In 127 of the 137 Trinitario individuals analysed, 51 alleles were detected using 25 RFLP probes. At 24 of these 25 loci, most Trinitario clones (127/137) had alleles observed in only two completely homozygous genotypes, one from Criollo (the most common genotype showed by 96% of 92 Ancient Criollo individuals studied by Motamayor *et al* (2002), and the other corresponding to the Lower Amazon Amelonado clone MAT1-6. Probe gtcCIR113 revealed three alleles within the Trinitario cultivar; alleles 1 or 2 were present in all Amelonado clones studied (MAT1-6 was homozygous for allele 2), and allele 3 was homozygous in Criollo individuals. In total, 49% of Trinitario clones (67/137) showed alleles 2 and/or 3. Thus, taking all loci studied into account, 49% of Trinitario individuals showed alleles present in MAT1-

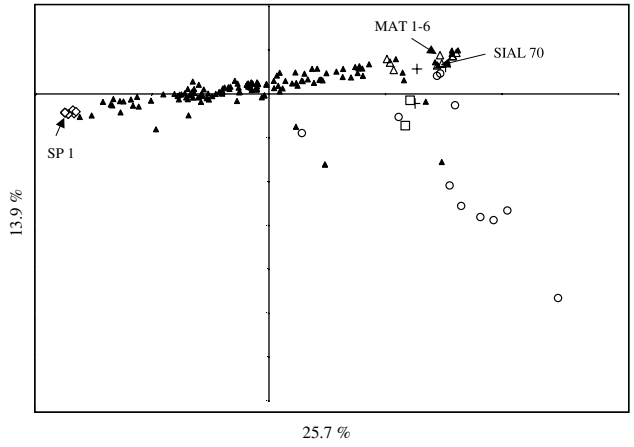

Figure 1 FAC from RFLP data: ◇, Criollo; ▲, Trinitario; △, Lower Amazon Amelonado; +, Orinoco Amelonado; □, French Guiana Amelonado; ○, hybrids with an Upper Amazon parent.

6 and in a typical Criollo. Of the remaining 51% of individuals, 54/137 had allele 1, found in Lower Amazon Amelonado clones such as SIAL70 and/or allele 3, from the most frequent Criollo genotype. Six Trinitario individuals were heterozygous for alleles 1 and 2 and therefore lacked the common Criollo allele. The other 10/137 Trinitario clones had alleles on six different loci specific to Upper Amazon Forastero individuals from Peru (Motamayor *et al*, 2002). Five of the six introgressed alleles in these individuals (CS05, MOQ 216, CRI 12, PV 02, COL9, CC 39, IFC 5, SNK625, JS 202/2, JS 211/21) were found in clone SCA9 and one was found in clone IMC67.

The RFLP data were used to carry out a FAC. This allows the genotypes to be represented according to the presence or absence of each RFLP allele. The first plane of the FAC (Figure 1) covers 39.6% of the total variability. In Figure 1, it can be observed that most Trinitario individuals are distributed along one continuous line between the Criollo and the Amelonado genotypes from the Lower Amazon, except those clones that showed introgression of alleles specific to the Upper Amazon region (Motamayor *et al*, 2002). Thus, Trinitario diversity consists of a combination of alleles from one Criollo (eg SP 1) and two Lower Amazon Amelonado genotypes (eg MAT1-6 and SIAL70). In Figure 1, it can be observed that individuals from the Orinoco and French Guiana tend to be placed outside the continuous Trinitario line delimited by the Lower Amazon Amelonado Forastero at the right-hand end of the line (second quadrant). Hybrids with at least one Upper Amazon Forastero parent showed high diversity.

Microsatellite results

To corroborate the RFLP results, a paternity analysis, including Amelonado individuals from Brazil, the Orinoco and French Guiana, was performed on a completely heterozygous Trinitario clone (ICS100). To increase the possible number of alleles detected per locus within the Trinitario cultivar, 16 microsatellites were used, which revealed a mean number of alleles per locus of 8.69 within a group of 28 Forastero individuals studied by Motamayor *et al* (2002). Table 2 shows the LOD score values obtained for the Amelonado candidate parents of clone ICS100, calculated with the most frequent Criollo genotype (represented by SP1) as the known parent. The most likely parent was MAT1-6, with a confidence level of 95%. The other two Lower Amazon Amelonado individuals SIC864 and SIAL70 had higher LOD scores than Amelonado from either Orinoco or French Guiana. Genotypes from the Orinoco had positive and negative LOD scores. Individuals VEN4 and VEN11 were collected in a gallery forest close to where a colonial plantation had existed (Lanaud, 1986); thus, it is possible that some gene flow may have occurred between wild Amelonado and cultivated Trinitario trees in this area. The standard likelihood ratios for parentage analysis were extended in Cervus 1.0 to take account of typing error (Marshall *et al*, 1998). If the error rate is greater than zero, no candidate parent is ever excluded, but those that mismatch at many loci acquire very low likelihood ratios (Marshall *et al*, 1998). Mismatches generated by alleles considered by Cervus 1.0 as null alleles are treated as if they were typing errors. As Amelonado and Criollo are fixed populations, Cervus 1.0 found high frequencies of null alleles from the allele frequency data (Wahlund effect), which explains why VEN11 and VEN20 had positive LOD scores. Mismatches were observed for eight microsatellite loci out of 16 for each of VEN4 and VEN11 regarding ICS100 and for nine microsatellites for VEN20. Therefore, performing paternity analysis through the excluding method (Chakraborty *et al*, 1974), individuals from the Orinoco were excluded as genotypes representing the putative parents (results not shown).

Since RFLP and parentage analyses showed that Lower Amazon Amelonado appeared to be involved in the origin of most Trinitario genotypes, the proportion of shared microsatellite alleles between Lower Amazon genotypes and Trinitario, as well as between Criollo and Trinitario, was analysed (Table 3). In total, 19 of the 36 Trinitario analysed have alleles that can be observed in two single individuals: SP1 (the Criollo showing the most frequent genotype) and a single Amelonado, MAT1-6. This result concurred with the paternity analysis of ICS100 and with the RFLP analysis where all the alleles scored for 67 of the 137 individuals studied were contained in the same two individuals (SP1 and MAT1-6). In total, 15 of the other 36 individuals (Table 3) shared microsatellite alleles with SP1 and genotypes closely related to MAT1-6 (P_{AS} between the Lower Amazon Amelonado genotypes higher than 0.81). Only CRI5 showed a higher P_{AS} value with a Criollo genotype slightly different from SP1 (LAN24, P_{AS} SP1-LAN24 = 0.938). Clone CRI5 also showed one allele found in Amelonado individuals from the Orinoco such as VEN11. CRI12 showed introgressions of Upper Amazon alleles at three loci as observed using RFLP markers. All the nine Trinitario individuals from Madagascar shared alleles with the most frequent Criollo genotype and either MAT1-6, SIAL70 or SIC19; SIC19 is a clone showing a genotype combining MAT1-6 and SIAL70 alleles. As mentioned above, individuals from Madagascar were analysed as the reference Trinitario.

Only Trinitario with P_{AS} values of 0.5 are the direct descendants of crosses involving homozygous Criollo and homozygous Amelonado genotypes, while indivi-

Table 2 LOD scores for the Amelonado candidate parents of clone ICS100, calculated considering the Criollo genotype SP1 as the known parent

	MAT1-6	*SIC864*	*SIAL70*	*VEN11*	*VEN20*	*VEN4*	*GU346*	*GU349*
LOD	10.16	6.24	5.07	1.28	1.10	−0.95	−4.01	−4.01
Delta	3.92							
Confidence level	*							

*Confidence level of 95%.

Table 3 Proportion of microsatellite alleles shared between the Trinitario individuals and the Criollo and Lower Amazon Amelonado genotypes that explain Trinitario's allelic constitution.

	LAN24	SP1	MAT1-6	SIC864	SIAL70	FSC15	SIC19	Σ
	(Criollo)		*(Lower Amazon Amelonado)*					
LAN24	1.000							
SP1	0.938	1.000						
MAT1-6	0.000	0.000	1.000					
SIC864	0.000	0.000	0.875	1.000				
SIAL70	0.000	0.000	0.813	0.875	1.000			
FSC15	0.000	0.000	0.938	0.938	0.844	1.000		
SIC19	0.000	0.000	0.906	0.875	0.906	0.938	1.000	
M589		0.300	0.700					1.000
M597		0.375	0.625					1.000
CRI37		0.469	0.531					1.000
CH120		0.594	0.406					1.000
CH42		0.500	0.500					1.000
ECH2		0.531	0.469					1.000
ICS40		0.533	0.467					1.000
ICS89		0.500	0.500					1.000
ICS95		0.594	0.406					1.000
ICS100		0.500	0.500					1.000
NOV29		0.667	0.333					1.000
PEN16		0.375	0.625					1.000
RIM113		0.500	0.500					1.000
RIM189		0.600	0.400					1.000
RIM68		0.375	0.625					1.000
SNK10		0.219	0.781					1.000
UF168		0.500	0.500					1.000
UF667		0.500	0.500					1.000
UF676		0.500	0.500					1.000
STA1		0.594	0.406					1.000
STA2		0.333	0.667					1.000
ATLER 1		0.267	0.733					1.000
M588		0.700			0.300			1.000
M591		0.600			0.400			1.000
M592		0.600			0.400			1.000
M594		0.600			0.400			1.000
DR1		0.344				0.656		1.000
IFC11		0.188				0.813		1.000
IFC19		0.219				0.781		1.000
ICS16		0.313				0.688		1.000
ICS53		0.500				0.500		1.000
ICS60		0.531					0.469	1.000
M596		0.400					0.600	1.000
M590		0.700					0.300	1.000
M593		0.367					0.633	1.000
CRI216		0.233					0.767	1.000
IFC4		0.156					0.844	1.000
IFC5		0.156					0.844	1.000
CRI 5	0.375		0.594					0.969
CRI 12		0.781	0.125					0.906

Amelonado individuals with Criollo introgression are shown in italics. FSC15: individual showing a genotype equivalent to a genotype resulting from the cross MAT1-6 × SIAL70; SIC19: individual showing a genotype equivalent to a genotype resulting from the cross MAT1-6 × SIC864. Σ: sum of the proportion of shared microsatellite alleles.

duals with different P_{AS} values indicate Trinitario intercrosses or selfing.

Another interesting fact that can be observed in Table 3 is that four Lower Amazon Amelonado types showed introgressions of Criollo microsatellite alleles (Table 3; IFC4, STA1, STA2 and Atelier1). This concurs with the RFLP results, where other Lower Amazon Forastero individuals from Honduras and Côte d'Ivoire also showed introgression of Criollo RFLP alleles (P_{AS} with SP1: TJ01 = 0.13; IFC2 = 0.03; IFC1 = 0.06; IFC15 = 0.12).

Diversity statistics for the Amelonado individuals from Brazil were calculated using microsatellite data. The genetic diversity for these individuals ($H_{n.b.} = 0.13$), the observed heterozygosity ($H_{obs} = 0.10$), the average number of alleles per locus ($A = 1.4$) and the percentage of polymorphic loci ($P_{0.95} = 0.31$) are low when compared with Forastero individuals from Upper Amazon countries for example Peru ($H_{n.b.} = 0.70$; $H_{obs} = 0.48$; $A = 5.5$, $P_{0.95} = 1$), using the same microsatellites on 13 individuals from this country (Motamayor *et al*, 2002). Within the Amelonado individuals from Brazil, 14 genotypes were found among the 17 individuals studied (Figure 2). Low D_{AS} values were found between nonidentical Amelonado Forastero clones (mean D_{AS} between clones under node A: 0.12, SD = 0.05), which explains the low bootstrap values obtained. It can be observed in Figure 2 that individual MAT1-6 from Costa Rica clusters with Amelonado individuals from Brazil.

Discussion

The historical data mentioning Criollo individuals as one of the parents of the Trinitario cultivar and the very narrow genetic diversity of Criollo observed by Motamayor *et al* (2002) prompted the search for the non-Criollo parent of Trinitario individuals.

In total, 49% of Trinitario individuals analysed through RFLP shared alleles with individuals SP1 (Criollo) and MAT1-6 (Amelonado) as did 53% when analysed through microsatellite markers. Similar results were obtained using both techniques despite differences in the size and constitution of the samples. Motamayor *et al* (2002) found a strong and significant correlation between two matrices of genetic distances (Pearson correlation coefficient = 0.9, probability of dependence = 1) obtained using the same microsatellite and RFLP markers employed in this work.

Considering the three genotypes SP1, MAT1-6 and SIAL70, 93% of the Trinitario individuals analysed showed only RFLP alleles observed in these genotypes. The same situation was observed for 86% of the Trinitario individuals analysed using microsatellites. In fact, the genotypes of the clones FSC15 and SIC19 are equivalent to the genotypes of individuals resulting from the crosses MAT1-6 × SIC864 and MAT1-6 × SIAL70, respectively. A third microsatellite allele, not observed in MAT1-6 or SIAL70, was detected in three Trinitario clones and two Amelonado clones SIC864 and SIC801. Therefore, the utilisation of microsatellite markers did not significantly change the number of genotypes implicated in the origin of most Trinitario individuals. For both techniques, around 50% of the Trinitario individuals analysed showed the allele diversity observed in only two individuals—SP1 and MAT1-6—and more than 75% in only three individuals: SP1, MAT1-6 and SIAL70. Paternity analysis clearly identified MAT1-6 as the most likely Amelonado parent of a completely heterozygous Trinitario clone, ICS100. Clone MAT1-6 was collected in Costa Rica; however, it is closely related to Brazilian clones FSC15 and SIC02 (Figure 2), indicating its Lower Amazon origin. Soria (1970) previously suggested the Lower Amazon origin of Amelonado

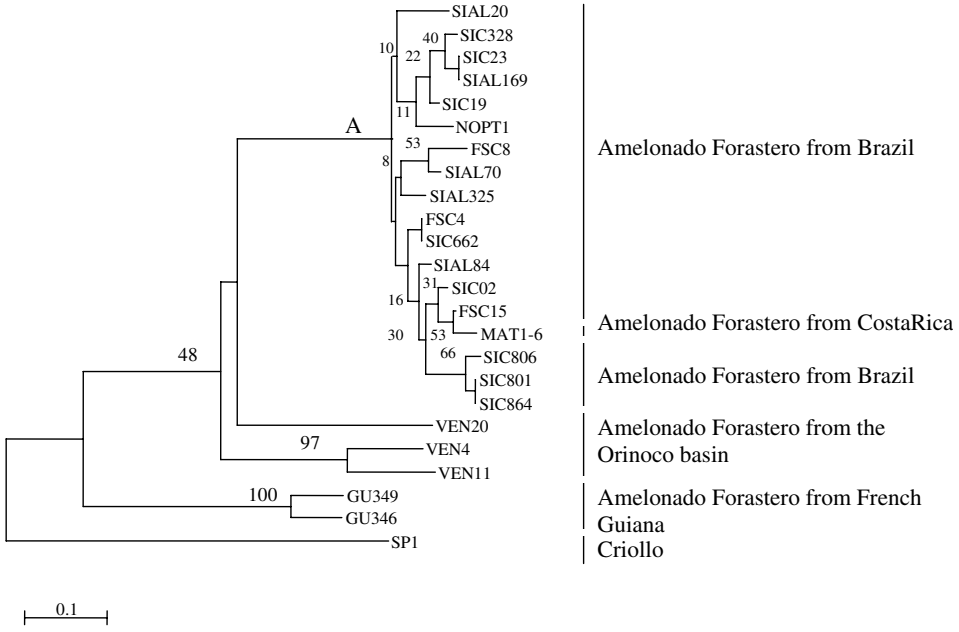

Figure 2 Neighbour-joining tree of Lower Amazon Forastero genotypes from Brazil and Costa Rica (MAT1-6) and of Amelonado Forastero individuals from the Orinoco basin and French Guiana, based on the shared allele distance calculated from microsatellite data. Criollo individual SP1 was included as outgroup. Bootstrap values were computed over 2000 replications by resampling loci and noted as percentages above each branch.

individuals from the Matina region, and their resemblance to Amelonado from Bahia (Brazil).

The number of analysed individuals from diverse geographic origins, and the occurrence of the alleles observed in SP1, MAT1-6 and SIAL70 in the reference group of Trinitario individuals from Madagascar, allowed an extrapolation of results to the entire cultivar. Thus, the geographic origin of Amelonado Forastero involved in the origin of Trinitario is the Lower Amazon. Furthermore, the results show the reduced genetic diversity of the Forastero individuals involved in the origin of Trinitario. This was unexpected, because as proposed by Warren (1994), given the close proximity of Trinidad (historically, the place where Trinitario originated) to the mainland, multiple introductions of different Forastero types seemed highly possible. In the analyses made by Motamayor *et al* (2002), each of the 37 Forastero individuals studied showed a different genotype. High genetic diversity within the Forastero group was previously reported (Lanaud, 1987; Laurent *et al*, 1994). Although a low genetic diversity was found for the Brazilian Amelonado individuals when compared with those Forastero from the Upper Amazon (eg from Peru, see above), the results suggest that a reduced number of Lower Amazon Amelonado genotypes was involved in the origin of Trinitario. In fact, five different alleles from four different loci observed in the Amelonado individuals from Brazil were absent in the Trinitario individuals analysed. These alleles contributed to the identification of the 14 genotypes observed for the 17 Amelonado individuals from Brazil analysed, from which only three genotypes, however, contain the allele diversity of the non-Criollo parent of Trinitario (MAT1-6, SIAL 70 and SIC864). In other words, only a subsample of the genetic diversity of the Amelonado individuals from Brazil is found in the Trinitario cultivar.

The relatively limited genetic basis of the Lower Amazon Amelonado individuals when compared with individuals from the Upper Amazon could be explained by historical data. Indeed, it has been stated that the Amelonado type cultivated in Bahia was introduced by one man, a Frenchman, Frederick Warneau, from Pará in 1746 (Wood, 1991), and it was from Bahia that most cacao would have been introduced into other regions such as Africa. The results presented here suggest that the Amelonado type introduced into Bahia could be the same as that introduced into Trinidad or Central America, since Trinitario and Amelonado trees from Bahia, Costa Rica, Nicaragua and Côte d'Ivoire share the same non-Criollo (Amelonado) alleles (Table 3). Introductions of the same Amelonado Forastero individuals could therefore be attributed to Spanish and Portuguese seafarers during the colonial period. Introductions probably started in Trinidad, and from there Amelonado may have been transported to Venezuela, Central America and the Caribbean. Amelonado individuals from the island of Santo Domingo are called Amelonado de Trinidad (Amelonado from Trinidad; Soria, 1970). In this way, Trinitario individuals could also have originated outside Trinidad, in the Criollo plantations where Amelonado individuals from the Lower Amazon were introduced. This could have happened in Central America for example, as suggested by Bartley (personal communications).

Since 1825, Trinitario individuals have replaced Criollo in most Latin American plantations (Pittier, 1935; Cheesman, 1944), because of their hybrid vigour and productivity. A similar situation has also been observed with the Nacional cultivar. This cultivar shows a narrow genetic base like the Criollo one (Lerceteau *et al*, 1997). Ecuadorian growers have gradually replaced Nacional trees by hybrids between Nacional and Trinitario individuals.

The Criollo and Amelonado genotypes described in this work, together with their hybrid, represent most cultivated cacao until 1950. As mentioned above, Criollo, Amelonado, Trinitario and Nacional (traditional cultivars) were the only cultivated cacaos until material collected by Pound (1938) in Peru and Ecuador became available to cacao-producing countries.

Cacao germplasm collections from different countries, for example Colombia, Côte d'Ivoire, Guatemala, Mexico, Nicaragua and Venezuela, maintain large numbers of Trinitario accessions, as do international germplasm collections in Trinidad and Costa Rica. Differences between Trinitario clones are mainly because of the proportion of alleles shared with the Criollo and the Amelonado original parents; thus, the maintenance of a high number of Trinitario accessions is not justified and numbers could be drastically reduced.

Since relatively few generations have elapsed since the formation of the Trinitario cultivar (considering a generation span of 20–30 years), linkage disequilibrium between specific Amelonado and Criollo agronomic traits and markers may have been maintained. This situation could be of great benefit in genome analyses of this cultivar, facilitating germplasm screening with molecular markers close to useful genes.

Acknowledgements

This study was partially financed by the International Plant Genetic Resources Institute (IPGRI) through the 1998 Vavilov-Frankel Fellowship, by the Consejo Nacional de Ciencia y Tecnologia (CONICIT) of Venezuela and by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, France).

References

Chakraborty R, Shaw M, Schull WJ (1974). Exclusion of paternity: the current state of the art. *Am J Hum Genet* **26**: 477–488.

Cheesman EE (1944). Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop Agric* **21**: 144–159.

Cuatrecasas J (1964). Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contrib US Herbarium* **35**: 379–614.

GENETIX Version. 4.0 *Logiciel sous Windows TM pour la génétique des populations*, Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.

Lanaud C (1986). *Rapport de prospection au Venezuela. Février 1986. IRCC-CIRAD*.

Lanaud C (1987). Nouvelles données sur la biologie du cacaoyer (Theobroma cacao L.): diversités de populations, systèmes d'incompatibilité, haploïdes spontanés; leurs conséquences pour l'amélioration génétique de cette espèce. PhD Dissertation, Université Paris XI.

Lanaud C, Risterucci AM, N'goran AKJ, Clément D, Flament MH, Laurent V *et al* (1995). A genetic linkage map of *Theobroma cacao* L. *Theor Appl Genet* **91**: 987–993.

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJL (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol Ecol* **8**: 2141–2152.

Laurent V, Risterucci AM, Lanaud C (1994). Genetic diversity in cocoa revealed by cDNA probes. *Theor Appl Genet* **88**: 193–198.

Lerceteau E, Robert T, Pétiard V, Crouzillat D (1997). Evaluation of the extent of genetic variability among *Theobroma cacao* L. accessions using RAPD and RFLP markers. *Theor Appl Genet* **95**: 10–19.

Lockwood G (1985). Genetic resources of the cocoa plant. *Span* **28**: 14–16.

Marshall TC, Slate J, Kruuk L, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* **7**: 639–655.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* **89**: 380–386.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Paulin D, Eskes AB (1995). Le cacaoyer: stratégies de sélection. *Plantations, Recherche. Développement* **2**: 5–18.

Pinto GC, Magalhaes WS, Silva E, Rodrigues EM (1967). Selection of high yielding cacao mother trees in Bahia. *Proceedings of the Second International Cacao Research Conference*. pp 105–112. Bahia, Brazil.

Pittier H (1935). Degeneration of cacao throught natural hybridization. *J Hered* **36**: 385–390.

Pound FJ (1938). *Cacao and Witchbroom disease (Marasmius perniciosus) of South America with Notes on Other Species of Theobroma*. Yuille's Printery: Port of Spain, Trinidad and Tobago (reprinted in 1982 in *Arch Cocoa Res*. **1**: 20–72).

Pound FJ (1945). A note about the cacao populations of South America. *Report and Proceedings Cocoa Research Conference, London*. Colonial 192:95–7. (reprinted 1982 in *Arch. Cocoa Res* **1**: 93–97).

Risterucci AM, Grivet L, N'goran JAK, Pieretti I, Flament MH and Lanaud C (2000). A high-density linkage map of *Theobroma cacao* L. *Theor Appl Genet* **101**: 948–955.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Silva P, Oliveira-Cardoso A (1980). Historico das introduçoes de cacaueiro (Theobroma cacao L.) no Recôncavo da Bahia, Brasil. *Revista Theobroma* **10**: 135–140.

Soria J (1970). Principal varieties of cacao cultivated in tropical America. *Cacao Growers' Bull* **15**: 12–21.

Toxopeus H (1985). Planting material. In: Wood GAR, Lass RA (eds) *Cacao*, 4th edn, Longman: London. pp 18–35.

Warren JM (1994). Isozyme variation in a number of populations of *Theobroma cacao* L. obtained through various sampling regimes. *Euphytica*, **72**: 121–126.

Wood GAR (1991). History of early cocoa introductions. *Cocoa Growers' Bull* **44**: 7–12.